Chapter 11

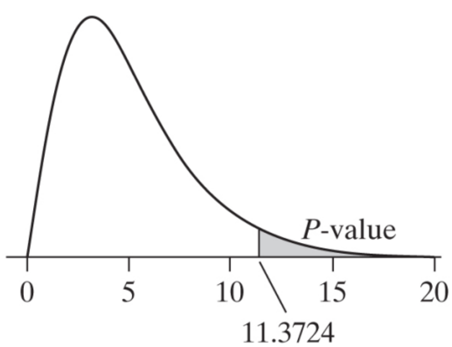
Section 11.1

Check Your Understanding, page 684:

1. H_0: The company's claimed color distribution for its Peanut M&M'S is correct versus H_a: The company's claimed color distribution is not correct. Or,
$H_0: p_{\text{blue}} = 0.23, p_{\text{orange}} = 0.23, p_{\text{green}} = 0.15, p_{\text{yellow}} = 0.15, p_{\text{red}} = 0.12, p_{\text{brown}} = 0.12$ versus
H_a: At least two of the p_i's is incorrect.
2. There were $12+7+13+4+8+2=46$ candies in the bag. The expected count of both blue and orange candies is $46(0.23)=10.58$, for green and yellow is $46(0.15)=6.9$, and for red and brown is $46(0.12)=5.52$.
3. $$\chi^2 = \frac{(12-10.58)^2}{10.58} + \frac{(7-10.58)^2}{10.58} + \frac{(13-6.9)^2}{6.9} + \frac{(4-6.9)^2}{6.9} + \frac{(8-5.52)^2}{5.52} + \frac{(2-5.52)^2}{5.52}$$
$$= 0.1906 + 1.2114 + 5.3928 + 1.2188 + 1.1142 + 2.2446 = 11.3724.$$

Check Your Understanding, page 687:

1. The expected counts (10.58, 10.58, 6.9. 6.9, 5.52, 5.52) are all at least 5. We should use the chi-square distribution with df = 6 – 1 = 5.
2.



Chi-square distribution with 5 df

3. From Table C, the *P*-value is between 0.025 and 0.05. From the calculator, *P*-value = χ^2cdf(lower: 11.3724, upper: 1000, df: 5) = 0.0445.
4. Because the *P*-value of 0.0445 is less than α = 0.05, we reject H_0. There is convincing evidence that the color distribution of M&M's Peanut Chocolate Candies is different from what the company claims.

Check Your Understanding, page 691:

1. *State:* We want to perform a test at the $\alpha = 0.01$ significance level of H_0: the distribution of eye color and wing shape is the same as what the biologists predict versus H_a: the distribution of eye color and wing shape is not the same as what the biologists predict. Or,
$H_0: p_{\text{red-straight}} = \frac{9}{16}, p_{\text{red-curly}} = \frac{3}{16}, p_{\text{white-straight}} = \frac{3}{16}, p_{\text{white-curly}} = \frac{1}{16}$ versus
H_a: At least two of the p_i's is incorrect. *Plan*: We should use a chi-square test for goodness of fit if the conditions are met. Random: The data are from a random sample. 10%: $n = 200$ is less than 10% of all fruit flies. Large Counts: The expected counts in each category are all at least 5 (red-straight: $200\left(\frac{9}{16}\right) = 112.5$, red-curly: $200\left(\frac{3}{16}\right) = 37.5$, white straight: $200\left(\frac{3}{16}\right) = 37.5$, and white-curly: $200\left(\frac{1}{16}\right) = 12.5$). *Do*: The test statistic is
$\chi^2 = \frac{(99-112.5)^2}{112.5} + \frac{(42-37.5)^2}{37.5} + \frac{(49-37.5)^2}{37.5} + \frac{(10-12.5)^2}{12.5} = 6.1867$. With df = 4 – 1 = 3, the P-value is between 0.10 and 0.15. *Using technology*: P-value = 0.1029. *Conclude*: Because the P-value of 0.1029 is greater than $\alpha = 0.01$, we fail to reject H_0. We do not have convincing evidence that the distribution of eye color and wing shape is different from what the biologists predict.

Exercises, page 693:

11.1 (a) H_0: The company's claimed distribution for its deluxe mixed nuts is correct versus H_a: The company's claimed distribution is not correct. Or,
$H_0: p_{\text{cashews}} = 0.52, p_{\text{almonds}} = 0.27, p_{\text{macadamia}} = 0.13, p_{\text{brazil}} = 0.08$ versus
H_a: At least two of the p_i's is incorrect.
(b) The expected count for cashews is $150(0.52) = 78$, for almonds is $150(0.27) = 40.5$, for macadamia nuts is $150(0.13) = 19.5$, and for brazil nuts is $150(0.08) = 12$.

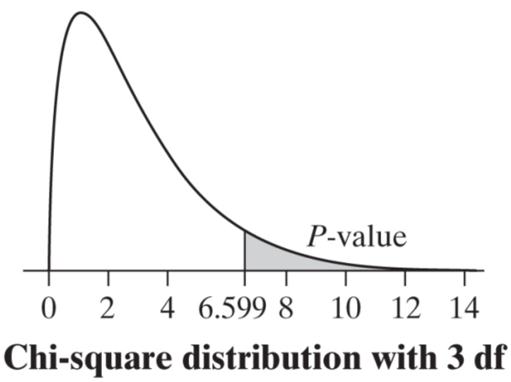
11.2 (a) H_0: The distribution of outcomes is what it should be versus H_a: The distribution of outcomes is not what it should be. Or, $H_0: p_{\text{red}} = \frac{18}{38}, p_{\text{black}} = \frac{18}{38}, p_{\text{green}} = \frac{2}{38}$ versus
H_a: At least two of the p_i's is incorrect.
(b) The expected count for red is $200\left(\frac{18}{38}\right) = 94.74$, for black is $200\left(\frac{18}{38}\right) = 94.74$, and for green is $200\left(\frac{2}{38}\right) = 10.53$.

11.3
$$\chi^2 = \frac{(83-78)^2}{78} + \frac{(29-40.5)^2}{40.5} + \frac{(20-19.5)^2}{19.5} + \frac{(18-12)^2}{12} = 0.321 + 3.265 + 0.013 + 3.000 = 6.599.$$

11.4 $\chi^2 = \frac{(85-94.74)^2}{94.74} + \frac{(99-94.74)^2}{94.74} + \frac{(16-10.53)^2}{10.53} = 1.001 + 0.192 + 2.841 = 4.034.$

11.5 (a) The expected counts calculated in Exercise 1 are all at least 5. Because there are 4 categories, use a chi-square distribution with df = 4 – 1 = 3.

(b)

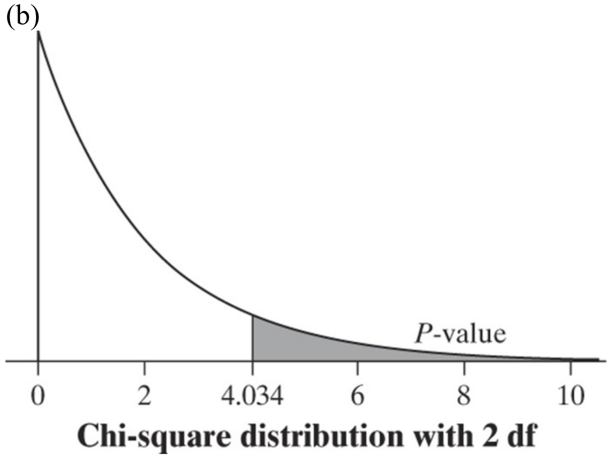


(c) Using Table C, the *P*-value is between 0.05 and 0.10. Using the calculator *P*-value = χ^2 cdf(lower: 6.599, upper: 1000, df: 3) = 0.0858.

(d) Because the *P*-value of 0.0858 is greater than α = 0.05, we fail to reject H_0. We do not have convincing evidence that the company's claimed distribution for its deluxe mixed nuts is incorrect.

11.6 (a) The expected counts calculated in Exercise 2 are all at least 5. Because there are 3 categories, use a chi-square distribution with df = 3 – 1 = 2.

(b)



(c) Using Table C, the *P*-value is between 0.10 and 0.15. Using the calculator, the *P*-value = χ^2 cdf(lower: 4.034, upper: 1000, df: 2) = 0.133.

(d) Because the *P*-value of 0.133 is greater than α = 0.05, we fail to reject H_0. We do not have convincing evidence that the distribution of outcomes on the roulette wheel is not what it should be.

11.7 *State:* We want to perform a test at the $\alpha = 0.05$ significance level of H_0: Nuthatches do not prefer particular types of trees when searching for seeds and insects versus H_a: Nuthatches do prefer particular types of trees when searching for seeds and insects; Or, $H_0: p_{\text{firs}} = 0.54, p_{\text{pines}} = 0.40, p_{\text{other}} = 0.06$ versus H_a: At least two of the p_i's is incorrect. *Plan:* We should use a chi-square test for goodness of fit if the conditions are met. Random: The data come from a random sample. 10%: n = 156 is less than 10% of all nuthatches. Large Counts: The expected counts in each category are all at least 5 (firs: $156(0.54) = 84.24$, pines: $156(0.40) = 62.4$, and other: $156(0.06) = 9.36$) *Do:* The test statistic is $\chi^2 = \frac{(70-84.24)^2}{84.24} + \frac{(79-62.4)^2}{62.4} + \frac{(7-9.36)^2}{9.36} = 7.418$. With df = 3 – 1 = 2, the *P*-value is between 0.02 and 0.025. *Using technology*: *P*-value = 0.0245. *Conclude:* Because the *P*-value of 0.0245 is less than α = 0.05, we reject H_0. There is convincing evidence that nuthatches prefer particular types of trees when they are searching for seeds and insects.

11.8 *State:* We want to perform a test at the $\alpha = 0.05$ significance level of H_0: Seagulls do not have a preference for where they land versus H_a: Seagulls do have a preference for where they land. Or, $H_0: p_{\text{sand}} = 0.56, p_{\text{mud}} = 0.29, p_{\text{rocks}} = 0.15$ versus H_a: At least two of the p_i's is incorrect. *Plan:* We should use a chi-square test for goodness of fit if the conditions are met. Random: The data come from a random sample. 10%: n = 200 is less than 10% of all seagulls. Large Counts: The expected counts in each category are at least 5 (sand: $200(0.56) = 112$, mud: $200(0.29) = 58$, rocks: $200(0.15) = 30$). *Do:* The test statistic is $\chi^2 = \frac{(128-112)^2}{112} + \frac{(61-58)^2}{58} + \frac{(11-30)^2}{30} = 14.474$. With df = 3 – 1 = 2, the *P*-value is between 0.0005 and 0.001. *Using technology: P*-value = 0.0007. *Conclude:* Because the *P*-value of 0.0007 is less than α = 0.05, we reject H_0. There is convincing evidence that seagulls have a preference for the location where they land.

11.9 A chi-square test for goodness of fit would not be appropriate here because time spent doing homework is quantitative. Chi-square tests for goodness of fit should only be used for distributions of categorical data.

11.10 A chi-square test for goodness of fit would not be appropriate here because the data do not describe the distribution of a single categorical variable. Homework completion status (yes or no) is being recorded for each student in the sample on all five days of the week.

11.11 (a) *State:* We want to perform a test at the α = 0.05 significance level of H_0: The first digit of invoices from this company follow Benford's law versus H_a: The first digit of invoices from this company do not follow Benford's law. Or, $H_0: p_1 = 0.301, p_2 = 0.176, p_3 = 0.125, p_4 = 0.097, p_5 = 0.079, p_6 = 0.067, p_7 = 0.058, p_8 = 0.051, p_9 = 0.046$ versus H_a: At least two of the p_i's is incorrect. *Plan:* We should use a chi-square test for goodness of fit if the conditions are met. Random: The data come from a random sample. 10%: We assume that n = 250 is less than 10% of all invoices from this company. Large Counts: The expected counts in each category are at least 5 (first digit 1: $250(0.301) = 75.25$, first digit 2: $250(0.176) = 44$, first digit 3: $250(0.125) = 31.25$, first digit 4: $250(0.097) = 24.25$, first digit 5: $250(0.079) = 19.75$, first digit 6: $250(0.067) = 16.75$, first digit 7: $250(0.058) = 14.5$, first digit 8: $250(0.051) = 12.75$, and first digit 9: $250(0.046) = 11.5$). *Do:* The test statistic is

$$\chi^2 = \frac{(61-75.25)^2}{75.25} + \frac{(50-44)^2}{44} + \frac{(43-31.25)^2}{31.25} + \frac{(34-24.25)^2}{24.25} + \frac{(25-19.75)^2}{19.75} + \frac{(16-16.75)^2}{16.75} + \frac{(7-14.5)^2}{14.5} + \frac{(8-12.75)^2}{12.75} + \frac{(6-11.5)^2}{11.5} = 21.563.$$

With df = 9 – 1 = 8, the P-value is between 0.005 and 0.01. *Using technology*: P-value = 0.0058. *Conclude:* Because the P-value of 0.0058 is less than α = 0.05, we reject H_0. There is convincing evidence that the first digit of invoices from this company do not follow Benford's law. *Follow-up analysis*: The breakdown of the chi-square statistic is: $\chi^2 = 2.699 + 0.818 + 4.418 + 3.920 + 1.396 + 0.034 + 3.879 + 1.770 + 2.630.$ From this we see that the largest contributors to the statistic are amounts with first digit 3, 4 and 7. There are more invoices that start with 3 or 4 than expected and fewer invoices that start with 7 than expected.

(b) A Type I error would be to find convincing evidence that the company's invoices do not follow Benford's law (suggesting fraud), when in reality they are consistent with Benford's law. A consequence is falsely accusing this company of fraud. A Type II error would be to not find convincing evidence that the invoices do not follow Benford's law (suggesting fraud), when in reality they do not. A consequence is allowing this company to continue committing fraud. A Type I error would be more serious for the accountant – alleging that the company had committed fraud when it had not – because he or she could be sued.

11.12 *State:* We want to perform a test at the $\alpha = 0.05$ significance level of H_0: The distribution of race in the complex is the same as the population distribution versus H_a: The distribution of race in the complex is not the same as the population distribution. Or, $H_0: p_{\text{Hispanic}} = 0.28, p_{\text{Black}} = 0.24, p_{\text{White}} = 0.35, p_{\text{Asian}} = 0.12, p_{\text{Others}} = 0.01$ versus H_a: At least two of the p_i's is incorrect. *Plan:* We should use a chi-square test for goodness of fit if the conditions are met. Random: The data come from a random sample. 10%: We assume that n = 800 is less than 10% of the residents in a large housing complex. Large Counts: The expected counts in each category are at least 5 (Hispanic: $800(0.28) = 224$, Black: $800(0.24) = 192$, White: $800(0.35) = 280$, Asian: $800(0.12) = 96$, and Others: $800(0.01) = 8$). *Do:* The test statistic is

$$\chi^2 = \frac{(212-224)^2}{224} + \frac{(202-192)^2}{192} + \frac{(270-280)^2}{280} + \frac{(94-96)^2}{96} + \frac{(22-8)^2}{8} = 26.0625.$$

With df = 5 – 1 = 4, the *P*-value is less than 0.0005. *Using technology*: *P*-value = 0.00003. *Conclude:* Because the *P*-value of 0.00003 is less than α = 0.05, we reject H_0. There is convincing evidence that the distribution of race in the complex is not the same as the population distribution. *Follow-up analysis*: The breakdown of the chi-square statistic is: $\chi^2 = 0.6429 + 0.5208 + 0.3571 + 0.0417 + 24.5$. From this we see that the largest contributor to the statistic, by far, is the other category. There were many more people in the "Other" category than we expected.

11.13 (a) H_0: The true distribution of flavors for Skittles candies is the same as the company's claim versus H_a: The true distribution of flavors for Skittles candies is not the same as the company's claim. Or, $H_0: p_{\text{lemon}} = 0.2, p_{\text{lime}} = 0.2, p_{\text{orange}} = 0.2, p_{\text{strawberry}} = 0.2, p_{\text{grape}} = 0.2$ versus H_a: At least two of the p_i's is incorrect.

(b) Because all 5 flavors have the same expected proportion, they all have the same expected count: $60(0.2) = 12$.

(c) Using df = 5 – 1 = 4 and Table C, the value for $\alpha = 0.05$ is 9.49 and for $\alpha = 0.01$ is 13.28. So, χ^2 statistics greater than 9.49 would provide significant evidence at the α = 0.05 level and χ^2 values greater than 13.28 would provide significant evidence at the α = 0.01 level.

(d) Answers will vary. One possibility consists of 6 lemon, 6 lime, 16 orange, 16 strawberry and 16 grape. The chi-square statistic is

$$\chi^2 = \frac{(6-12)^2}{12} + \frac{(6-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(16-12)^2}{12} = 10,$$

which is between 9.49 and 13.28.

11.14 (a) H_0: Each of the digits 0–9 from my random number generator is equally likely versus H_a: Each of the digits 0–9 from my random number generator is not equally likely. Or, $H_0: p_1 = 0.1, p_2 = 0.1, p_3 = 0.1, p_4 = 0.1, p_5 = 0.1, p_6 = 0.1, p_7 = 0.1, p_8 = 0.1, p_9 = 0.1, p_0 = 0.1$ versus H_a: At least two of the p_i's is incorrect.

(b) Answers will vary. We perform the test at the $\alpha = 0.05$ significance level. *Plan:* We should use a chi-square test for goodness of fit if the conditions are met. Random: The data come from a random sample. 10%: Not needed, because we are not sampling without replacement. Large Counts: The expected count in each category is at least 5 (all digits: 200(0.10) = 20). *Do:* In one test of the calculator we got 18 0's, 22 1's, 23 2's, 21 3's, 21 4's, 21 5's, 17 6's, 14 7's, 21 8's, and 22 9's. The test statistic is

$$\chi^2 = \frac{(18-20)^2}{20} + \frac{(22-20)^2}{20} + \frac{(23-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(17-20)^2}{20} + \frac{(14-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(22-20)^2}{20} = 3.5$$

With df = 10 – 1 = 9, the *P*-value is greater than 0.25. *Using technology*: *P*-value = 0.9411. *Conclude:* Because the *P*-value of 0.9411 is greater than α = 0.05, we fail to reject H_0. We don't have convincing evidence that each of the digits 0–9 from my random number generator is not equally likely.
(c) Because we are using α = 0.05, there is a 0.05 probability of committing a Type I error, assuming that the random number generator is working properly.
(d) *P*(at least one Type I error) = 1 – *P*(no Type I errors) = 1 – $(0.95)^{25}$ = 1 – 0.277 = 0.723. There is a 0.723 probability that at least one student makes a Type I error, assuming that their random number generators are all working properly.

11.15 *State:* We want to perform a test of H_0: All 12 astrological signs are equally likely versus H_a: All 12 astrological signs are not equally likely. Or, $H_0: p_i = \frac{1}{12}$ versus H_a: At least two of the p_i's is incorrect, where each p_i is the true proportion of people with each astrological sign. We will use α = 0.05. *Plan:* We should use a chi-square test for goodness of fit if the conditions are met. Random: The data come from a random sample. 10%: n = 4344 is less than 10% of all people in the United States. Large Counts: The expected counts in each category are all at least 5 (because the total sample size is 4344 and there are 12 months, the expected count is $4344\left(\frac{1}{12}\right) = 362$ for each month). *Do:* The test statistic is

$$\chi^2 = \frac{(321-362)^2}{362} + \frac{(360-362)^2}{362} + \ldots + \frac{(355-362)^2}{362} = 19.76.$$

With df = 12 – 1 = 11, the *P*-value is between 0.025 and 0.05. *Using technology*: *P*-value = 0.0487. *Conclude:* Because the *P*-value of 0.0487 is less than α = 0.05, we reject H_0. There is convincing evidence that the 12 astrological signs are not equally likely.
Follow-up analysis: The breakdown of the chi-square statistic is: χ^2 = 4.64 + 0.01 + 0.07 + 0.40 + 1.22 + 4.42 + 2.49 + 3.01 + 2.65 + 0.18 + 0.54 + 0.14. From this we see that the largest contributors to the statistic are Aries and Virgo. There are fewer Aries (321 – 362 = –41) and more Virgos (402 – 362 = 40) than we would expect.

11.16 *State:* We want to perform a test at the $\alpha = 0.05$ significance level of H_0: Froot Loops contain an equal proportion of each flavor versus H_a: Froot Loops do not contain an equal proportion of each flavor. Or, $H_0: p_i = \frac{1}{6}$ versus H_a: At least two of the p_i's is incorrect, where each p_i is the true proportion of Froot Loops of each flavor. *Plan:* We should use a chi-square test for goodness of fit if the conditions are met. Random: The data come from a random sample. 10%: $n = 120$ is less than 10% of all Froot Loops. Large Counts: The expected counts in each category are all at least 5 (because the total sample size is 120, the expected count in each category is $120\left(\frac{1}{6}\right) = 20$). *Do:* The test statistic is

$$\chi^2 = \frac{(28-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(16-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(14-20)^2}{20} + \frac{(16-20)^2}{20} = 7.9.$$

With df = 6 – 1 = 5, the *P*-value is between 0.15 and 0.20. *Using technology*: *P*-value = 0.1618. *Conclude:* Because the *P*-value of 0.1618 is greater than α = 0.05, we fail to reject H_0. We do not have convincing evidence that Froot Loops do not contain an equal proportion of each flavor. *Follow-up analysis*: Not necessary.

11.17 *State:* We want to perform a test at the $\alpha = 0.05$ significance level of H_0: Mendel's 3:1 genetic model is correct versus H_a: Mendel's 3:1 genetic model is not correct. Or, $H_0: p_{\text{smooth}} = 0.75, p_{\text{wrinkled}} = 0.25$ versus H_a: Both of the proportions are incorrect. *Plan:* We should use a chi-square test for goodness of fit if the conditions are met. We are told to assume the conditions for inference are met. *Do:* The test statistic is

$$\chi^2 = \frac{(423-417)^2}{417} + \frac{(133-139)^2}{139} = 0.3453.$$

With df = 2 – 1 = 1, the *P*-value is greater than 0.25. *Using technology*: *P*-value = 0.5568. *Conclude:* Because the *P*-value of 0.5568 is greater than α = 0.05, we fail to reject H_0. We do not have convincing evidence that Mendel's 3:1 genetic model is not correct.

11.18 *State:* We want to perform a test at the $\alpha = 0.05$ significance level of H_0: The proposed 9:3:3:1 genetic model is correct versus H_a: The proposed 9:3:3:1 genetic model is not correct. Or, $H_0: p_{\text{tall-cut}} = 0.5625, p_{\text{tall-potato}} = 0.1875, p_{\text{dwarf-cut}} = 0.1875, p_{\text{dwarf-potato}} = 0.0625$ versus H_a: At least two of the proportions is incorrect. *Plan:* We should use a chi-square test for goodness of fit if the conditions are met. We are told to assume the conditions for inference are met. *Do:* The test statistic is

$$\chi^2 = \frac{(926-906.1875)^2}{906.1875} + \frac{(288-302.0625)^2}{302.0625} + \frac{(293-302.0625)^2}{302.0625} + \frac{(104-100.6875)^2}{100.6875} = 1.469.$$

Using df = 4 – 1 = 3, the *P*-value is greater than 0.25. *Using technology*: *P*-value = 0.6895. *Conclude:* Because the *P*-value of 0.6895 is greater than α = 0.05, we fail to reject H_0. We do not have convincing evidence that the proposed 9:3:3:1 genetic model is not correct.

11.19 d

11.20 a

11.21 c

11.22 c

11.23 The distribution of English grades for the heavy readers is skewed to the left while the distribution of English grades for the light readers is roughly symmetric. The center of the distribution of English grades is greater for the heavy readers than for the light readers, indicating that heavy readers typically get higher grades in English. The English grades are more variable for the light readers. There is one low outlier in the heavy reading group but no outliers in the light reading group.

11.24 (a) The conditions are met. Random: Even though the data came from a single random sample, it is reasonable to consider the two samples independent because knowing the response of a heavy reader shouldn't help us predict the response of a light reader. 10%: $n_1 = 47$ is less than 10% of heavy readers and $n_2 = 32$ is less than 10% of light readers at this large school. Normal/Large Sample: $n_1 = 47 \geq 30$ and $n_2 = 32 \geq 30$.
(b) *State:* Our parameters of interest are $\mu_1 =$ the true mean English grade of heavy readers and $\mu_2 =$ the true mean English grade of light readers. We want to estimate the difference $\mu_1 - \mu_2$ at a 95% confidence level. *Plan:* We should use a two-sample *t* interval for $\mu_1 - \mu_2$ if the conditions are met. We checked the conditions in part (a). *Do:* The conservative degrees of freedom is 32 – 1 = 31. Using Table B and df = 30, the 95% confidence interval is $(3.64 - 3.356) \pm 2.042\sqrt{\frac{(0.324)^2}{47} + \frac{(0.38)^2}{32}} = 0.284 \pm 0.1677 = (0.1163, 0.4517)$. Using technology: (0.1197, 0.4483) with df = 59.46. *Conclude:* We are 95% confident that the interval from 0.1197 to 0.4483 captures the true difference in the mean English grade of heavy and light readers.
(c) No. Even though 0 is not in the confidence interval, this was an observational study so no conclusion of cause and effect can be made.

11.25 (a) The slope is 0.024. For each additional book read, the predicted English GPA increases by about 0.024. The *y* intercept is 3.42. The predicted English grade for a student who has read 0 books is about 3.42.
(b) The predicted English GPA for this student is $\hat{y} = 3.42 + 0.024(17) = 3.828$. The residual is $y - \hat{y} = 2.85 - 3.828 = -0.978$. This student's English GPA is 0.978 less than predicted, based on the number of books this student has read.
(c) The relationship between English grades and number of books read is not very strong. On the scatterplot, the points are quite spread out from the line. Also, the value of r^2 is only 0.083, which means that only 8.3% of the variation in English grades is accounted for by the linear model relating English GPA to number of books read.

11.26 (a) What Luis has calculated is the probability of getting 5 of one particular number (e.g. getting all sixes). What he has not taken into account is that there are 6 different possible ways of getting a Yahtzee – one for each number on the die. (b) No, Nassir should not be surprised. The probability of getting a Yahtzee on any roll is $6\left(\frac{1}{6}\right)^5 = 0.000772$. The probability of getting no Yahtzee on any roll is $1 - 0.000772 = 0.999228$. Finally, the probability of getting 25 rolls with no Yahtzee is very large: $(0.999228)^{25} = 0.9809$.